



LAS CITAS RECIBIDAS COMO PRINCIPAL FACTOR DE POSICIONAMIENTO SEO EN LA ORDENACIÓN DE RESULTADOS DE GOOGLE SCHOLAR

Received citations as a main SEO factor of Google Scholar results ranking



Cristòfol Rovira, Frederic Guerrero-Solé y Lluís Codina

Note: This article can be read in its original English version on:
<http://www.elprofesionaldelainformacion.com/contenidos/2018/may/09.pdf>



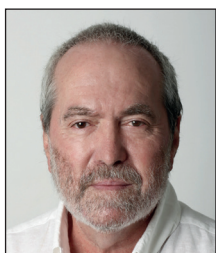
Cristòfol Rovira es profesor de la *Universitat Pompeu Fabra (UPF)* desde 1992. Imparte docencia en los grados de Periodismo y de Publicidad. Es director del *Máster Universitario en Documentación Digital*, del *Máster Universitario en Buscadores* y el *Máster Universitario en Usabilidad (UPF-Barcelona School of Management)*. Es licenciado en Ciencias de la Educación, ingeniero técnico en Informática de Gestión, diplomado en Biblioteconomía y Documentación y master en Software libre. Investiga en posicionamiento web (SEO), usabilidad, marketing en buscadores y mapas conceptuales con técnicas de *eyetracking*.
<https://orcid.org/0000-0002-6463-3216>

cristofol.rovira@upf.edu



Frederic Guerrero-Solé es licenciado en Física por la *Universitat de Barcelona (UB)*, y doctor en Comunicación Pública por la *Universitat Pompeu Fabra (UPF)*. Imparte docencia en la *Facultad de Comunicación* de la *UPF* desde 2008, donde es profesor de Sociología de la Comunicación. Es miembro del grupo de investigación *Unidad de Investigación en Comunicación Audiovisual (Unica)*.
<https://orcid.org/0000-0001-8145-8707>

frederic.guerrero@upf.edu



Lluís Codina es profesor titular del *Departamento de Comunicación* de la *Universitat Pompeu Fabra (UPF)*, donde ha impartido docencia en los grados de Periodismo y de Comunicación Audiovisual durante los últimos 25 años. Actualmente es el coordinador del *Master Universitario en Comunicación Social*. Es doctor en Ciencias de la Información por la *Universitat Autònoma de Barcelona*, en la que fue investigador de su programa de doctorado y profesor ayudante. Imparte docencia en el *Master Online en Documentación* y en el de *Buscadores* en la *UPF-Barcelona School of Management*.
<https://orcid.org/0000-0001-7020-1631>

lluis.codina@upf.edu

*Universitat Pompeu Fabra, Departament de Comunicació
Roc Boronat, 138. 08018 Barcelona, España*

Resumen

El objetivo de este artículo es analizar los factores de posicionamiento (SEO) externos que pueden influir en la ordenación por relevancia en *Google Scholar* y luego identificar el peso las citas recibidas en esta ordenación. Se ha aplicado una metodología de ingeniería inversa comparando el ranking de *Google Scholar* con un ranking formado tan sólo por el número de citas recibidas por los documentos. El estudio se realizó a partir de cuatro tipos de búsquedas sin palabras clave: por publicación, año, autor y "citado por". Los resultados fueron coincidentes en las cuatro muestras con coeficientes de correlación entre los dos rankings superiores al 0,9. El presente estudio demuestra de forma más clara que en anteriores investigaciones que las citas recibidas es el factor SEO externo más relevante en el ranking de los resultados en *Google Scholar*. Los demás factores tienen una influencia mínima. Esta información proporciona una base sólida para la disciplina del posicionamiento

en buscadores académicos (ASEO). También hemos desarrollado una nueva propuesta metodológica que aísla los factores SEO externos y que puede ser útil en futuras investigaciones.

Palabras clave

ASEO; SEO; Ingeniería inversa; Citaciones; *Google Scholar*; Indicadores; Rankings; Algoritmos; Ordenación; Motores de búsqueda académicos.

Summary

The aim of this article is to analyze the web positioning factors that can influence the order, by relevance, in *Google Scholar* and the subsequent evaluation of the importance of received citations in this ordering process. The methodology of reverse engineering was applied, in which a comparison was made between the *Google Scholar* ranking and another ranking consisting of only the number of citations received by documents. This investigation was conducted employing four types of searches without the use of keywords: by publication, year, author, and "cited by". The results were matched in the four samples with correlation coefficients between the two highest rankings, which exceeded 0.9. The present study demonstrates more clearly than in previous research how citations are the most relevant off-page feature in the ranking of search results on *Google Scholar*. The other features have minimal influence. This information provides a solid basis for the academic search engine optimization (ASEO) discipline. We also developed a new analysis procedure for isolating off-page features that might be of practical use in forthcoming investigations.

Keywords

ASEO; SEO; Reverse engineering; Citations; *Google Scholar*; Indicators; Rankings; Algorithms; Academic search engines.

Rovira, Cristòfol; Guerrero-Solé, Frederic; Codina, Lluís (2018). "Received citations as a main SEO factor of *Google Scholar* results ranking". *El profesional de la información*, v. 27, n. 3, pp. 559-569.

<https://doi.org/10.3145/epi.2018.may.09>

1. Introducción

La optimización de motores de búsqueda (SEO) es el proceso empleado para optimizar los sitios web y su contenido para tener una buena posición en los listados de resultados de los buscadores (Engel; Spencer; Stricchiola, 2015). SEO también es una profesión bien establecida dentro de la nueva industria de la comunicación digital como demuestra la existencia de empresas especializadas en su práctica, editores profesionales y formación universitaria. Su objetivo es destacar la calidad de los documentos para aumentar su visibilidad en relación con los algoritmos que establecen las posiciones en los motores de búsqueda, especialmente *Google*. Este objetivo debe lograrse sin falsificar las características de los documentos y sin emplear medios fraudulentos.

La razón dada por *Google* para esta falta de transparencia es luchar contra el spam

Las páginas de resultados de búsqueda de *Google* están ordenadas por relevancia (*Google*, 2017), que se basa en más de 200 características, aunque *Google* no especifica cuáles son o su peso específico; simplemente proporciona información parcial y general, como por ejemplo que la calidad del contenido y los enlaces entrantes son dos factores predominantes (Ratcliff, 2016; Schwartz, 2016).

El motivo indicado por *Google* para esta falta de transparencia es luchar contra el spam (Beel; Gipp, 2010). Si todos los detalles de los factores de ordenación estuvieran disponibles, documentos de baja calidad se podrían colocar en buenas posiciones. Sin embargo esta política de caja negra

va en detrimento de los profesionales de SEO que llevan a cabo su actividad de una manera ética y cuyo trabajo se ve obstaculizado por la falta de información fiable.

Algunas empresas de SEO (Giesen; Rosen, 2016; Local SEO guide, 2016; MOZ, 2015; Searchmetrics, 2016) realizan investigaciones de ingeniería inversa para medir el impacto de los factores involucrados en el proceso de posicionamiento de *Google*. En esas investigaciones se realizan muchas búsquedas para identificar los factores de posicionamiento en función de las características de las páginas ubicadas en las primeras posiciones. Dado que la cantidad de factores que intervienen en el proceso de posicionamiento es grande, es extremadamente difícil establecer qué factores son realmente relevantes y en qué medida influyen en la ordenación final de los documentos. Además el proceso de posicionamiento en *Google* es altamente dinámico, ya que el algoritmo sufre docenas de cambios por año (MOZ, 2017).

El ASEO (*academic search engine optimisation*) es el posicionamiento aplicado a los buscadores académicos

En los últimos años el SEO se ha aplicado a los motores de búsqueda académicos, denominándose este nuevo proceso como SEO académico (ASEO) (Beel; Gipp, 2009b, 2010; Codina, 2016; Martín-Martín et al., 2016a; Muñoz-Martín, 2015). Los autores ponen cada vez más énfasis en mejorar la visibilidad de sus artículos en los buscadores académicos. Los artículos que aparecen en las primeras posiciones aumentan su visibilidad y por tanto tienen una mayor probabilidad de ser leídos y citados, lo que a su vez aumenta las probabilidades de mejorar el índice h personal de sus autores (Farhadi et al., 2013).

Register for free at <https://www.scipedia.com> to download the version without the watermark

En muchos casos los mismos procedimientos de optimización utilizados con éxito en *Google* se están aplicando a *Google Scholar*. Sin embargo este último tiene su propio algoritmo. Aunque pocos estudios abordan los factores de ordenación específicos de *Google Scholar*, algunos que pueden citarse son: **Beel y Gipp** (2009b; 2009c; 2010); **Beel, Gipp y Wilde** (2010); **Martín-Martín et al.** (2014; 2017); **Orduña-Malea et al.** (2016).

El objetivo del presente estudio es analizar las características de los documentos que pueden influir en la ordenación por relevancia en *Google Scholar*, y particularmente nos interesan las citas recibidas por los artículos. El objetivo es evaluar la influencia del número de citas recibidas en el algoritmo de ordenación. La cantidad de veces que se cita un documento es una característica clave para determinar la especificidad del proceso de ordenación de *Google Scholar*. Pensamos que la influencia de las citas es mucho mayor de lo que generalmente piensan autores y editores. Por ejemplo, las instrucciones para los autores de las publicaciones académicas proporcionan sugerencias sobre cómo mejorar la posición de los artículos en el ranking en *Google Scholar* (*Elsevier*, 2012; *Wiley*, 2015; *Emerald Publishing Limited*, 2017). En estas guías las citas recibidas no se mencionan o se tratan sin la importancia que tienen.

Este artículo presenta los resultados de un estudio de ingeniería inversa que utiliza un nuevo método de análisis. Este método permite bloquear algunos factores del algoritmo de posicionamiento, en concreto aquellos que dependen de elementos externos de la página ordenada. De este modo se puede enfocar el estudio sobre un pequeño conjunto de factores que son estudiados con mayor control.

La hipótesis que se plantea es que si se compara la ordenación aplicando sólo el número de citas recibidas con la ordenación estándar (<https://www.scipedia.com>), donde sólo participan factores externos, se puede identificar el peso de las citas en el conjunto de estos factores externos. Si los dos rankings comparados son similares entonces las citas tendrán un peso importante.

Esta nueva metodología es posible gracias al formulario de búsqueda avanzada de *Google Scholar* que permite a los usuarios restringir los campos de búsqueda a autor, año y publicación. En estos tipos de búsqueda no hay palabras clave y sólo participan factores externos. De este modo los resultados que se han obtenido son mucho más fiables que en estudios previos que aplicaron la ingeniería inversa en *Google*, pero sin este control de variables.

2. Trabajos previos

Google Scholar se ha convertido en una alternativa a los servicios de indexación de citas científicas clásicas, como *Web of Science (WoS)* o *Scopus*. La posición de estos servicios de indexación en el mercado puede verse comprometida si *Google Scholar* ofrece un producto gratuito de calidad similar. Por esta razón, *Google Scholar* ha sido analizado desde diversas perspectivas:

- Análisis comparativo, longitudinal o de cobertura, intentando determinar su calidad y utilidad (**Giustini; Boulos**, 2013; **Walters**, 2008; **De-Winter; Zadpoor; Dodou**, 2014;

Harzing, 2013; 2014; **De-Groote; Raszewski**, 2012; **Orduña-Malea et al.**, 2014; 2015; **Pedersen; Arendt**, 2014; **Jamali; Nabavi**, 2015).

- Evaluación del impacto de los autores, de sus citaciones o del índice h (**Van-Aalst**, 2010; **Jacsó**, 2008a; 2008b; 2009; 2012; **Martín-Martín et al.**, 2014; 2017; **Farhadi et al.**, 2013).
- Evaluación de la utilidad de *Google Scholar* para estudios bibliométricos sobre la calidad de la actividad científica (**Aguillo**, 2012; **Jacsó**, 2009; **Torres-Salinas; Ruiz-Pérez; Delgado-López-Cózar**, 2009; **Beel; Gipp**, 2010; **Delgado-López-Cózar et al.**, 2012; 2014; **Martín-Martín et al.**, 2016b).

Sin embargo, se han llevado a cabo pocas investigaciones sobre el proceso de recuperación de la información, la efectividad de las búsquedas (**Jamali; Asadi**, 2010; **Walters**, 2008) y sobre los factores que intervienen en los algoritmos de ordenación por relevancia (**Beel; Gipp**, 2009a; 2009b; 2009c; **Beel; Gipp; Wilde**, 2010). A diferencia del *Google* generalista (*Google Search*), el estudio del proceso de posicionamiento en *Google Scholar* ha despertado poco interés científico, lo cual es sorprendente ya que influye en qué artículos se leerán más. Es ampliamente reconocido que los primeros elementos que aparecen en una lista de resultados de búsqueda reciben más atención de los usuarios (**Marcos; González-Caro**, 2010). Una mejor posición en el ranking implica mejores posibilidades de ser encontrado y leído.

A diferencia del proceso de posicionamiento en *Google Search*, el utilizado en *Google Scholar* ha despertado poco interés científico.

A partir de estos trabajos sobre la ordenación por relevancia y *Google Scholar*, se proponen varias hipótesis:

- Las palabras clave utilizadas en la búsqueda deben aparecer en el título para favorecer un buen posicionamiento del documento (**Beel; Gipp**, 2009a).
- La frecuencia de las palabras clave en el texto del documento no parece ser un factor determinante para establecer su ordenación (**Beel; Gipp**, 2009a).
- Los artículos recientes tienen una posición más baja que los artículos más antiguos (**Beel; Gipp**, 2009a) para compensar el “efecto Mateo” (**Merton**, 1968): los artículos con muchas citas se clasifican en los primeros puestos; por lo tanto, tienen más lectores y más citas y, en consecuencia, consolidan sus posiciones en la parte superior (**Martín-Martín et al.**, 2016b).
- El número de citas recibidas es un factor determinante para establecer la ordenación por relevancia (**Beel; Gipp**, 2009c; **Martín-Martín et al.**, 2014).

Esta última conclusión es particularmente relevante para el presente estudio. Sin embargo las investigaciones que la sustentan tienen algunas limitaciones. En **Beel y Gipp** (2009c) se analizaron todos los factores SEO al mismo tiempo. Por tanto las variables relacionadas con los factores internos no se bloquearon y los resultados no son lo suficientemente claros. En **Martín-Martín et al.** (2014) se usaron sólo muestras de búsquedas por año.

Register for free at <https://www.scipedia.com> to download the version without the watermark

El objetivo central de la presente investigación es corroborar esa conclusión mediante la aplicación de una metodología que establece un control más estricto sobre las variables. Esta metodología permite obtener información precisa sobre la relevancia de las citas recibidas en relación con todas las características externas del algoritmo de ordenación en *Google Scholar*.

3. Ingeniería inversa

La ingeniería inversa es un método de análisis para obtener información sobre cómo se diseñó un sistema o dispositivo. Generalmente se usa para estudiar dispositivos electrónicos e identificar sus componentes y la forma de funcionar. También se utiliza en la elaboración de software para obtener el código fuente de los programas compilados.

Ingeniería inversa en buscadores implica calcular los coeficientes de correlación entre la posición que ocupa una página y los valores de los factores que supuestamente intervienen en el algoritmo

La ingeniería inversa se ha aplicado a las búsquedas de *Google* para determinar los factores que intervienen en la ordenación por relevancia (Localseoguide, 2016; MOZ, 2015; Searchmetrics, 2016). A partir de la información parcial proporcionada por *Google* (2017), se estudian las páginas de resultados para identificar el funcionamiento de los algoritmos de ordenación. Se examinan las características de los documentos en las primeras posiciones para determinar los factores que intervienen y sus ponderaciones. La ingeniería inversa requiere un gran esfuerzo ya que los algoritmos son propietarios y están sujetos a *Non-disclosure Agreements* (Verder Graaf, 2012).

El proceso de ingeniería inversa en motores de búsqueda generalmente implica el cálculo de los coeficientes de correlación de Spearman entre la posición que ocupa una página en una búsqueda y los valores de los factores que supuestamente intervienen en el algoritmo de ordenación. Una correlación alta indica que se puede atribuir un mayor peso a la característica que se está analizando en su contribución a

la ordenación. En general, una correlación entre 0,4 y 0,7 se considera moderada, y una mayor de 0,7 se considera alta. Sin embargo, en casos más complejos, como el actual, que involucra un gran número de variables, los coeficientes de correlación raramente exceden de 0,3 (MOZ, 2015).

Aunque *Google* no proporciona información detallada sobre cómo funciona el algoritmo de ordenación, sí que facilita información general sobre qué factores intervienen. A partir de esta información parcial, se puede deducir que hay más de 200 factores involucrados en sus criterios de ordenación por relevancia, que se pueden dividir en dos tipos:

- factores internos: están relacionados con el contenido del documento y con la presencia de palabras clave utilizadas de las búsquedas en el texto de estos documentos (Engel; Spencer; Stricchiola, 2015);
- factores externos: son indicadores de calidad relacionados con el contexto del documento y no con su contenido; por ejemplo, la calidad y cantidad de enlaces de entrada calculados mediante el *PageRank* (Maciá-Domene, 2015).

Estudios previos aplicando la ingeniería inversa han encontrado grandes dificultades para analizar estas 200 variables (Localseoguide, 2016; MOZ, 2015; Searchmetrics, 2016).

La situación es aún peor en el caso de *Google Scholar* ya que la información proporcionada sobre los algoritmos de ordenación es aún más escasa. Una de las pocas explicaciones explícitas sobre su funcionamiento es que:

“*Google Scholar* tiene como objetivo ordenar los documentos de la misma forma que lo hacen los investigadores, ponderando el texto completo de cada documento, dónde se publicó, quién lo escribió, y con qué frecuencia y como de reciente ha sido citado en otra bibliografía académica” (Google, 2011).

A pesar de no disponer de los parámetros de los investigadores, se puede afirmar que el algoritmo de *Google Scholar* es más simple y con menos factores que el de *Google* generalista (*Google Search*) (Mayr; Walter, 2007; Torres-Salinas; Ruiz-Pérez; Delgado-López-Cózar, 2009). En *Google Scholar* no hay evidencia de la presencia de muchos de los factores que intervienen en *Google Search* (tabla 1).

Afortunadamente *Google Scholar* tiene varias funciones de búsqueda avanzada no disponibles en *Google Search* que

Tabla 1. Factores SEO de *Google Search* y *Google Scholar*

Tipo	Factor SEO	Google Search	Google Scholar
Factores internos	Relevancia del contenido: palabras clave en el título, URL, h1, primeras 100 palabras	Sí	Sí
	Factores técnicos: diseño adaptativo, velocidad de carga, usabilidad, metadatos y datos estructurados, HTTP.	Sí	?
Factores externos	Enlaces de entrada, <i>PageRank</i>	Sí	?
	Citas recibidas	No	Sí
	Reputación del autor	Sí	Sí
	Reputación de la publicación o del dominio	Sí	Sí
	Señales sociales	?	?
	Tráfico, CTR	Sí	?
Internos + externos	<i>RankBrain</i>	Sí	?
	<i>Machine-learning</i> , inteligencia artificial	Sí	?

permiten a los investigadores controlar los factores internos. Estos quedan deshabilitados cuando las búsquedas son realizadas por autor, editor, año, o por medio del enlace “citado por”. En estos casos la ordenación por relevancia se realiza sólo mediante la aplicación de factores externos. Las variables relacionadas con las características internas de la página no juegan ningún papel, ya que en la búsqueda no se utilizan las palabras clave. Son búsquedas binarias; por lo tanto, un documento puede o no ser de un autor, año o trabajo publicado.

La búsqueda por campos de *Google Scholar* permite estudiar el algoritmo de ordenación cuando sólo intervienen los factores SEO externos

Figura 1. Búsqueda avanzada de *Google Scholar*

Un segundo factor hace que la ingeniería inversa sea particularmente productiva. *Google Scholar* da información sobre los valores exactos de las citas recibidas de cada documento, uno de los principales factores externos de ordenación. Esta información es muy valiosa para la ingeniería inversa. Si comparamos el ranking estándar de *Google Scholar* con el orden que resulta de aplicar sólo la cantidad de veces que se cita cada documento, podremos obtener una buena aproximación del peso de las citas en los algoritmos de ordenación. Si ambos rankings son similares significará que las citas son un factor importante.

Si además podemos controlar los factores internos, entonces podremos obtener evidencias adicionales. El método estadístico que nos permite analizar estos datos consiste en calcular el coeficiente de correlación entre el ranking convencional de *Google Scholar* y el orden que se obtiene al aplicar sólo el número de citas.

únicamente en factores externos, en los que no participen los factores internos relacionados con la coincidencia de las palabras clave en la búsqueda y en el documento.

Para lograr este control de los factores internos, se han utilizado cuatro tipos de búsquedas básicas en las que las palabras clave temáticas no desempeñan ningún papel. Se han utilizado los campos de publicación, autor y año de la búsqueda avanzada de *Google Scholar* (figura 1). También empleamos un cuarto tipo de búsqueda no temática mediante el uso del enlace “citado por”, que está disponible en cada ítem de los listados de resultados de *Google Scholar* (figura 2). Este enlace permite realizar una nueva búsqueda y obtener los trabajos que citan el documento inicial que tienen el enlace.

Register for free at <https://www.scipedia.com> to download the version without the watermark

4. Metodología

En esta investigación comparamos el ranking de *Google Scholar* con otro ranking, que se creó al aplicar sólo el número de citas recibidas en los documentos de los listados de resultados. Por lo tanto, comparamos un ranking complejo que de acuerdo con *Google* (2011) tiene en cuenta al menos el texto completo, la publicación, el autor y las citas recibidas, con otro ranking mucho más simple formado por una sola de estas variables, las citas recibidas. Esta comparación puede darnos una aproximación del peso de las citas en el algoritmo de ordenación de *Google Scholar*. Si los rankings son casi iguales, entonces el número de citas será un factor importante en la ordenación.

Sin embargo, el número de citas es un factor externo. Por lo tanto, la comparación debe hacerse con rankings de *Google Scholar* basados

únicamente en factores externos, en los que no participen los factores internos relacionados con la coincidencia de las palabras clave en la búsqueda y en el documento.



Figura 2. Enlace “Citado por” en los listados de resultados de *Google Scholar*

blish or Perish extrae automáticamente los resultados de las búsquedas realizadas en *Google Scholar*. Es una de las pocas aplicaciones para hacer este tipo de procesos puesto que *Google* no permite la extracción masiva de datos.

En las búsquedas sin palabras clave en *Google Scholar*, el orden por el número de citas recibidas es casi idéntico al orden por relevancia

En cada uno de estos cuatro tipos de búsqueda, se realizaron 25 búsquedas de 1.000 resultados, alcanzando un total de 100.000 datos. En cada caso, se comparó el ranking de *Google Scholar* con el orden obtenido al considerar sólo el número de citas. Para el análisis estadístico se aplicó el coeficiente de correlación de Spearman porque las distribuciones no eran normales.

Las muestras fueron elegidas evitando sesgos. Las búsquedas excluyeron los documentos de patentes porque no siguen el mismo patrón de citas que los artículos académicos. También tratamos de asegurarnos de que el volumen de citas en todos los documentos recuperados fuera relativamente alto para evitar resultados con pocas citas recibidas. La ordenación de estos resultados no citados debe realizarse con otros factores, por lo tanto, los datos estarían contaminados. Los procedimientos de selección de las cuatro muestras fueron los siguientes:

Tabla 2. Coeficientes de correlación globales para cada tipo de búsqueda ($\alpha = 0,0025$)

Tipo de búsqueda	r de Spearman	L	U
Publicación	0,999	0,999	0,999
Año	0,909	0,898	0,919
Autor	0,998	0,997	0,998
Citado por	0,999	0,999	0,999
Global	0,999	0,999	0,999

- publicaciones: la selección de las publicaciones se realizó aleatoriamente;
- años: se seleccionó el período de 25 años entre 1989 y 2013. Los cuatro años entre 2014 y 2017 fueron excluidos para lograr un volumen similar de citas para todos los años;
- autores: la búsqueda fue realizada por apellidos. Se seleccionaron los apellidos más comunes en el Reino Unido y Estados Unidos ya que son autores muy comunes en *Google Scholar*. Cada búsqueda incluyó varios autores con el mismo apellido;
- citado por: la elección de los artículos se realizó aleatoriamente.

Los datos obtenidos de *Google Scholar* sobre el número de citas de cada documento de la lista de resultados se transformaron en rangos (escala ordinal) (Beel; Gipp, 2009c). Por lo tanto se creó una ordenación alternativa para cada búsqueda de acuerdo sólo con el número de citas, que luego se comparó con la ordenación por relevancia estándar de *Google Scholar* por medio del coeficiente de correlación de Spearman.

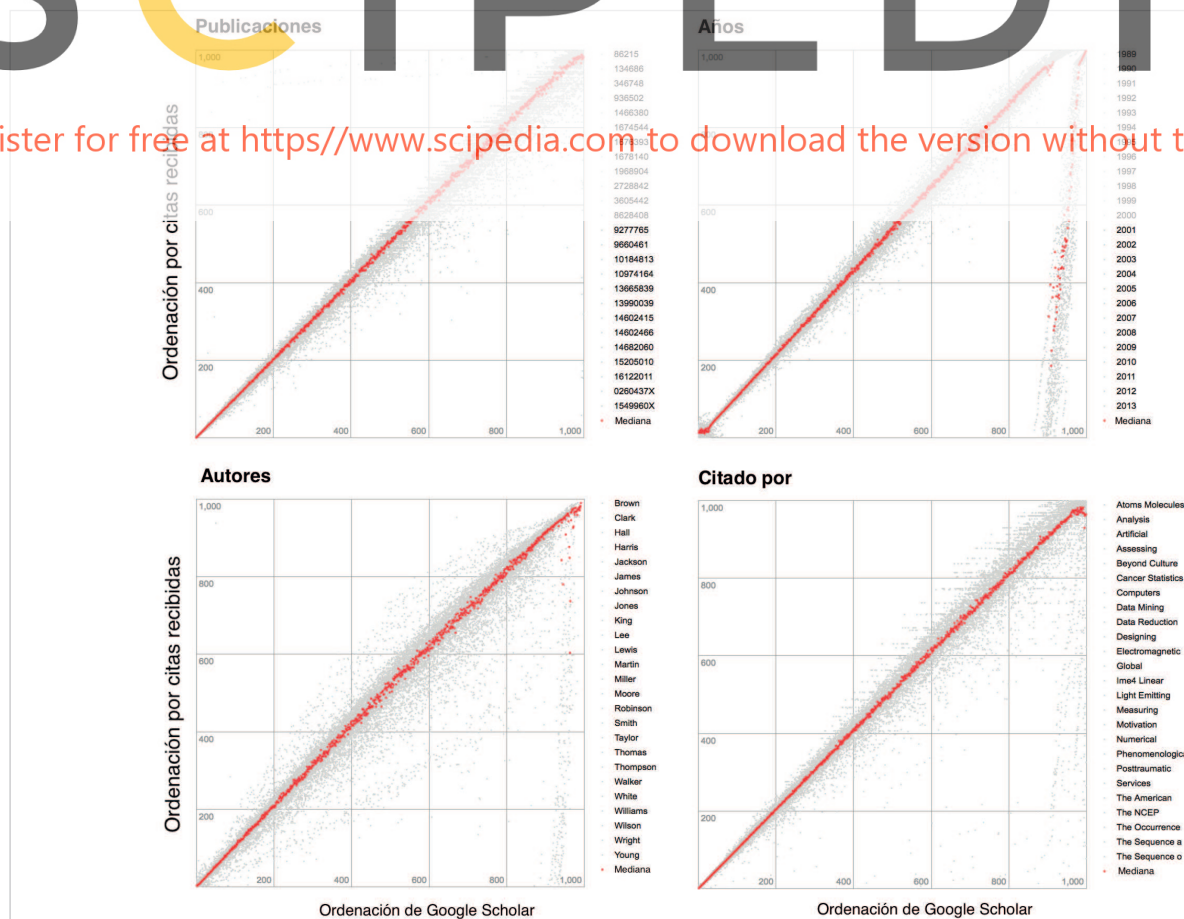


Figura 3. Diagrama de dispersión de las cuatro muestras correspondientes a los cuatro tipos de búsqueda

Para agregar los datos para cada tipo de búsqueda y obtener coeficientes de correlación generales, se comparó la posición en *Google Scholar* de cada documento con el promedio de las 25 muestras. También se aplicó el promedio para calcular el valor total de las cuatro muestras con el análisis de 100.000 datos.

El software utilizado en el análisis fue *R 3.4.0* (*R Development Core Team*, 2017). Los intervalos de confianza se han construido con la aproximación normal mediante la transformación de Fisher, utilizando el paquete de *R Psych* (Revelle, 2017). La transformación de Fisher aplicada a coeficientes de correlación de Spearman es asintóticamente normal. Los gráficos de los intervalos de confianza se han realizado con el paquete *Plotrix* de *R* (Lemon, 2006).

Para garantizar una confianza total del 95% en la interpretación de 25 intervalos de confianza, se ha utilizado un nivel de significación de $\alpha = 0,0025$ siguiendo el criterio conservador de Bonferroni. El hecho de exigir un nivel de significación bajo hace que los intervalos de confianza sean más anchos que con un valor α superior (p.ej., 0,05).

5. Resultados

Los coeficientes de correlación de Spearman entre el ranking de *Google Scholar* y los rankings de citas recibidas fueron sorprendentemente altos, con valores cercanos o superiores a 0,9 (tabla 2).

Las figuras 3 y 4 muestran que las correlaciones para las cuatro muestras fueron casi perfectas. Tan sólo desde la posición número 900 encontramos datos que no correlacionan, especialmente en el caso de la búsqueda por años. Es un efecto ya detectado en anteriores investigaciones (Martín-Martín *et al.*, 2014; 2017). Los datos obtenidos muestran que en estas posiciones el número de citas recibidas es muy bajo. Por tanto es muy posible que en estos casos la posición del ranking se establezca por otros factores SEO externos, distintos a las citas.

Si sólo se hubieran considerado los valores hasta la posición número 900, los coeficientes de correlación hubieran sido aún mayores. La tabla 3 presenta los datos específicos para cada búsqueda. Los coeficientes de correlación fueron todos cercanos a 0,9, y excedieron este valor en la mayoría de los casos.

6. Discusión

Los coeficientes de correlación fueron sorprendentemente altos. En aplicaciones anteriores de ingeniería inversa para el ranking de búsqueda de *Google*, raramente se obtuvieron coeficientes superiores a 0,4. Nuestra investigación ha obtenido coeficientes de correlación mayores a 0,9 para los cuatro tipos de búsqueda y un valor global de 0,9999. Aunque los cuatro tipos de búsqueda son completamente diferentes entre sí (autores, años, publicaciones y citado por), en todas

Tabla 3. Coeficientes de correlación de Spearman y límites de confianza de las cuatro muestras correspondientes a los cuatro tipos de búsqueda ($\alpha = 0,0025$)

#	Años				Autores				Citado por				Publicaciones			
	id	rho	L	U	id	rho	L	U	id	rho	L	U	id	rho	L	U
1	1989	0,90	0,88	0,91	Clark	0,97	0,96	0,97	Atoms_molecules	0,99	0,99	0,99	14682060	0,99	0,99	0,99
2	1990	0,87	0,84	0,89	Hall	0,91	0,89	0,93	Analysis	0,99	0,99	0,99	9660461	0,99	0,99	0,99
3	1991	0,87	0,85	0,89	Johnson	0,95	0,93	0,97	Artificial	0,99	0,99	0,99	14682060	0,99	0,99	0,99
4	1992	0,90	0,88	0,92	Jackson	0,93	0,91	0,94	Assessing	0,99	0,99	0,99	2728842	0,98	0,98	0,99
5	1993	0,89	0,87	0,91	James	0,97	0,96	0,98	Beyond_culture	0,94	0,93	0,95	15205010	0,99	0,99	0,99
6	1994	0,89	0,87	0,91	King	0,96	0,95	0,96	Cancer_statistics	0,99	0,99	0,99	9277765	0,99	0,99	0,99
7	1995	0,88	0,86	0,90	Lee	0,96	0,95	0,97	Computers	0,98	0,97	0,98	936502	0,99	0,99	0,99
8	1996	0,89	0,87	0,91	Lewis	0,97	0,96	0,97	Data_Mining	0,98	0,97	0,98	134686	0,99	0,99	0,99
9	1997	0,89	0,87	0,91	Martin	0,90	0,88	0,92	Data_reduction	0,99	0,99	0,99	1968904	0,99	0,99	0,99
10	1998	0,89	0,87	0,91	Moore	0,96	0,95	0,96	Designing	0,95	0,93	0,95	3605442	0,99	0,99	0,99
11	1999	0,87	0,85	0,89	Robinson	0,95	0,94	0,96	Electromagnetic	0,98	0,98	0,99	10184813	0,99	0,99	0,99
12	2000	0,89	0,87	0,91	Taylor	0,96	0,95	0,96	Global	0,98	0,97	0,98	8628408	0,99	0,99	0,99
13	2001	0,89	0,87	0,91	Thomas	0,90	0,87	0,91	Ime4_Linear	0,95	0,94	0,96	0260437X	0,99	0,99	0,99
14	2002	0,86	0,83	0,88	Thompson	0,94	0,93	0,95	Light_emitting	0,99	0,99	0,99	1674544	0,98	0,98	0,99
15	2003	0,88	0,86	0,90	Walker	0,96	0,95	0,97	Measuring	0,99	0,99	0,99	1549960X	0,99	0,99	0,99
16	2004	0,89	0,86	0,91	White	0,96	0,95	0,97	Motivation	0,90	0,88	0,92	14602466	0,99	0,99	0,99
17	2005	0,92	0,90	0,93	Wright	0,95	0,94	0,96	Numerical	0,97	0,96	0,97	13665839	0,99	0,98	0,99
18	2006	0,83	0,79	0,85	Young	0,94	0,93	0,95	Phenomenological	0,94	0,93	0,95	14602415	0,99	0,99	0,99
19	2007	0,84	0,81	0,86	Wilson	0,97	0,97	0,98	Posttraumatic	0,98	0,97	0,98	16122011	0,99	0,99	0,99
20	2008	0,88	0,86	0,90	Brown	0,97	0,96	0,97	Services	0,94	0,93	0,95	10974164	0,99	0,99	0,99
21	2009	0,90	0,88	0,91	Johnson	0,97	0,96	0,97	The_American	0,99	0,99	0,99	1466380	0,99	0,99	0,99
22	2010	0,88	0,85	0,90	Jones	0,97	0,96	0,97	The_NCEP	0,98	0,98	0,99	1678140	0,99	0,99	0,99
23	2011	0,83	0,80	0,86	Miller	0,97	0,97	0,98	The_occurrence	0,99	0,99	0,99	346748	0,88	0,85	0,90
24	2012	0,89	0,86	0,91	Smith	0,96	0,95	0,97	The_sequence a	0,90	0,88	0,92	1676393	0,99	0,98	0,99
25	2013	0,89	0,86	0,91	Williams	0,96	0,95	0,97	The_sequence o	0,99	0,99	0,99	13990039	0,99	0,99	0,99

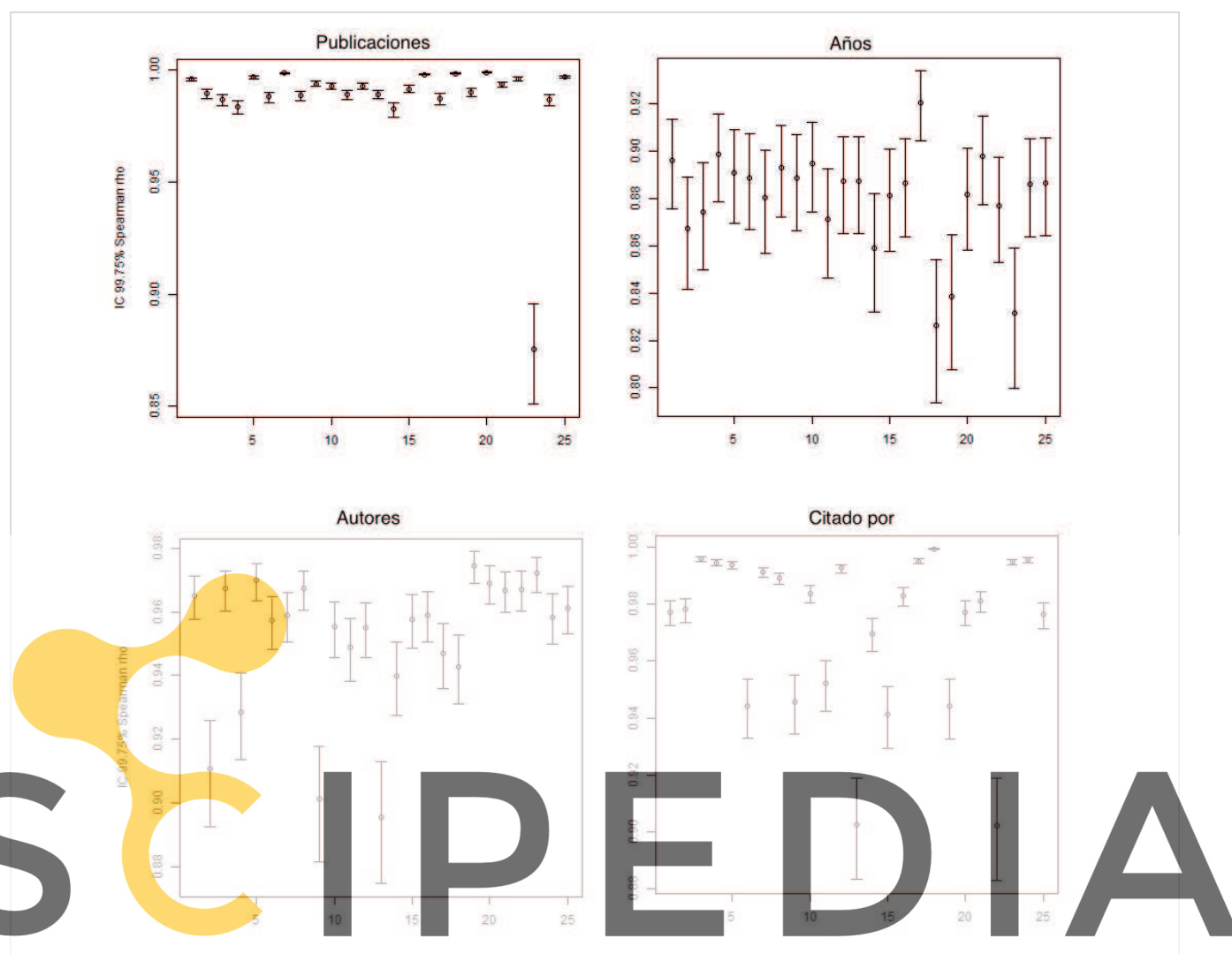


Figura 4. Diagrama de caja de las cuatro muestras correspondientes a los cuatro tipos de búsqueda

Register for free at <https://www.scipedia.com> to download the version without the watermark

ellas se obtuvo el mismo patrón de correlación. No obstante, las correlaciones por años presentan una mayor variabilidad que el resto de búsquedas y los autores a su vez una mayor variabilidad que publicaciones y citado por (figura 4).

Estos resultados ofrecen una evidencia concluyente de que el número de citas recibidas es de suma importancia cuando se consideran los factores SEO externos en *Google Scholar*. Estas conclusiones son coherentes con las de **Beel y Gipp** (2009a) y **Martín-Martín et al.** (2014; 2017), quienes abordaron el mismo problema desde un ángulo metodológico diferente:

- En el primer caso se analizaron búsquedas temáticas con palabras clave llegando a la misma conclusión de que las citas recibidas es un factor de ordenación de gran relevancia. No obstante, estos resultados no son lo suficientemente claros porque no se analizaron los factores SEO externos de forma aislada.
- En **Martín-Martín et al.** (2014; 2017) se estudiaron sólo las búsquedas por año obteniendo igualmente coeficientes de correlación del orden del 0,9.

Nuestro estudio aporta evidencias adicionales al analizar también las búsquedas por autores, publicaciones y citado por, obteniéndose en todos los casos las mismas altas correlaciones. Por otro lado, en **Martín-Martín et al.** (2017) se

correlaciona la posición en *Google Scholar* con el número de citas recibidas; por tanto se está correlacionando una variable ordinal con otra cuantitativa discreta. En cambio nuestra propuesta metodológica consiste en comparar dos variables ordinales, el orden de *Google Scholar* con el orden según las citas (**Beel; Gipp**, 2009a) a partir de la mediana de cada muestra.

Las publicaciones académicas dan instrucciones a los autores sobre cómo mejorar el ranking en GS sin dar a las citas recibidas la importancia que merecen

Algunas investigaciones (**Martín-Martín et al.**, 2017; **Moed; Bar-Ilan; Halevi**, 2016) apuntan la posibilidad de que otros factores externos como el idioma del documento, el número de versiones o la velocidad de indexación puedan influir en el ranking. No obstante, con correlaciones de 0,99 sobre un máximo de 1 queda muy poco margen para factores que no sean las citas. La ordenación por número de citas es prácticamente igual a la ordenación nativa de *Google Scholar*. Los factores externos restantes ejercen tan sólo una

influencia residual, incluidos los enlaces recibidos, que son muy importantes en la búsqueda de Google. Estos factores podrían tener un mayor protagonismo en documentos con pocas o ninguna cita recibida, como ha ocurrido en los ítems situados a partir de la posición 900.

Las guías o instrucciones de los editores de publicaciones académicas dirigidas a los autores generalmente dan pautas sobre cómo mejorar la posición de los artículos en el ranking en Google Scholar (Elsevier, 2012; Wiley, 2015). Estas pautas tienden a estar contaminadas por el ranking de Google. Por ejemplo, a menudo se afirma que el orden de ordenación en Google Scholar depende del editor y se insinúa que algunas publicaciones tienen un mejor posicionamiento. Las sugerencias a los autores también afirman que existen otros factores externos de posicionamiento, como la presencia en redes sociales, los enlaces de entrada o el prestigio del autor según el número de citas recibidas en todos sus trabajos publicados (Google, 2011).

Para que un artículo suba en el ranking de Google Scholar tiene que recibir citas

No se han encontrado evidencias de que estas afirmaciones correspondan a factores que intervengan en el algoritmo de ordenación de Google Scholar. Sin embargo, pueden ejercer una influencia indirecta si permiten que el trabajo publicado sea más leído y, por lo tanto, aumente el número de citas recibidas.

A partir de este estudio no podemos extraer conclusiones sobre los factores SEO internos relacionados con el uso de palabras clave en los documentos para mejorar el posicionamiento. Esta característica no se ha estudiado directamente puesto que se usaron búsquedas sin palabras clave y los factores internos estuvieron bloqueados. Por lo tanto, no hay evidencias en contra de las recomendaciones que generalmente se dan en este contexto, como por ejemplo incluir las palabras clave más importantes del artículo y sus sinónimos en el título, subtítulos o resumen y optimizar el número de veces que aparece una palabra clave en el artículo.

7. Conclusiones

Hemos desarrollado un nuevo procedimiento de análisis en el contexto de los estudios de ingeniería inversa que nos permitió estudiar de forma aislada los factores SEO externos en Google Scholar. A través de este nuevo método, hemos podido confirmar que las citas recibidas son el principal factor SEO externo en Google Scholar. Los resultados estadísticos no dejan lugar a dudas. Este análisis ayuda a aumentar las evidencias que se tenían sobre este tema y presenta un nuevo enfoque estadístico que puede aplicarse en otras investigaciones.

Al emplear ingeniería inversa, pudimos obtener estimaciones de los factores que intervienen en la ordenación de los resultados y su importancia relativa. Nuestros hallazgos son útiles para mejorar la base experimental de la disciplina del ASEO y para proporcionar mejores recomendaciones a los autores sobre cómo pueden optimizar el posicionamiento de sus trabajos en Google Scholar.

En conclusión, esta investigación además de demostrar un valor intrínseco para el ASEO, ofrece recomendaciones específicas para autores de artículos científicos y académicos. La recomendación principal que se puede extraer a partir de este estudio es hacer artículos de calidad para que sean leídos y tengan citas y de este modo entren en el “círculo virtuoso” de cuantas más citas mejor ranking y cuanto mejor ranking, más visibilidad y más citas (Martín-Martín et al., 2016b).

8. Agradecimientos

Este trabajo forma parte del proyecto *Creación y contenido interactivo en la comunicación de información audiovisual: audiencias, diseño, sistemas y formatos*, CSO2015-64955-C4-2-R Ministerio de Economía y Competitividad (Mineco/Feder), España.

9. Referencias

- Aguillo, Isidro F.** (2012). “Is Google Scholar useful for bibliometrics? A webometric analysis”. *Scientometrics*, v. 91, n. 2, pp. 343-351.
<https://goo.gl/nYBmZb>
- Beel, Joeran; Gipp, Bela** (2009a). “Google Scholar’s ranking algorithm: An introductory overview”. In: *Procs of the 12th intl conf on scientometrics and informetrics, ISSI’09*, pp. 230-241.
<https://goo.gl/c8a6YU>
- Beel, Joeran; Gipp, Bela** (2009b). “Google Scholar’s ranking algorithm: the impact of articles’ age (an empirical study)”. In: *6th intl conf on information technology: New generations, ITNG’09*, pp. 160-164.
<https://goo.gl/cfV2my>
<https://doi.org/10.1109/ITNG.2009.317>
- Beel, Joeran; Gipp, Bela** (2009c). “Google Scholar’s ranking algorithm: the impact of citation counts (an empirical study)”. In: *13th intl conf on scientometrics and informetrics, RCIS 2009*, pp. 439-446.
<https://www.gipp.com/wp-content/papercite-data/pdf/beel09a.pdf>
<https://doi.org/10.1109/RCIS.2009.5089308>
- Beel, Joeran; Gipp, Bela** (2010). “Academic search engine spam and Google Scholar’s resilience against it”. *The journal of electronic publishing*, v. 13, n. 3, pp. 1-28.
<https://doi.org/10.3998/3336451.0013.305>
- Beel, Joeran; Gipp, Bela; Wilde, Erik** (2010). “Academic search engine optimization (ASEO). Optimizing scholarly literature for Google Scholar & co”. *Journal of scholarly publishing*, v. 41, n. 2, pp. 176-190.
[https://docear.org/papers/Academic%20Search%20Engine%20Optimization%20\(ASEO\)%20--%20preprint.pdf](https://docear.org/papers/Academic%20Search%20Engine%20Optimization%20(ASEO)%20--%20preprint.pdf)
<https://doi.org/10.3138/jsp.41.2.176>
- Codina, Lluís** (2017). “SEO académico: definición, componentes y guía de herramientas”. *Lluís Codina*, 17 noviembre.
<https://www.lluiscodina.com/seo-academico-guia>
- De-Groote, Sandra L.; Raszewski, Rebecca** (2012). “Coverage of Google Scholar, Scopus, and Web of Science: A case study of the h-index in nursing”. *Nursing outlook*, v. 60, n. 6, pp. 391-400.
<https://doi.org/10.1016/j.outlook.2012.04.007>

Register for free at <https://www.scipedia.com> to download the version without the watermark

De-Winter, Joost C. F.; Zadpoor, Amir A.; Dodou, Dimitra (2014). "The expansion of Google Scholar versus Web of Science: A longitudinal study". *Scientometrics*, v. 98, n. 2, pp. 1547-1565.

<https://doi.org/10.1007/s11192-013-1089-2>

Delgado-López-Cózar, Emilio; Robinson-García, Nicolás; Torres-Salinas, Daniel (2012). *Manipular Google Scholar Citations y Google Scholar Metrics: simple, sencillo y tentador. EC3 working papers*. Granada: Universidad De Granada. <http://hdl.handle.net/10481/20469>

Delgado-López-Cózar, Emilio; Robinson-García, Nicolás; Torres-Salinas, Daniel (2014). "The Google Scholar experiment: How to index false papers and manipulate bibliometric indicators". *Journal of the Association for Information Science and Technology*, v. 65, n. 3, pp. 446-454. <https://arxiv.org/abs/1309.2413> <https://doi.org/10.1002/asi.23056>

Elsevier (2012). "Get found: optimize your research articles for search engines". Elsevier, 6 Nov. <https://www.elsevier.com/connect/get-found-optimize-your-research-articles-for-search-engines>

Emerald Publishing Limited (2017). "How to... disseminate your work". Emerald Publishing. <http://www.emeraldgrouppublishing.com/authors/guides/promote/disseminate.htm>

Eng, Eric; Spencer, Stephan; Stricchiola, Jessie (2015). *The art of SEO: mastering search engine optimization*. Sebastopol, CA: O'Reilly Media. ISBN: 978 1 491903643

Farhadi, Hadi; Salehi, Hadi; Yunus, Melor; Aghaei-Chadegani, Arezoo; Farhadi, Maryam; Fooladi, Masood; Ale-Ebrahim, Nader (2013). "Does it matter which citation tool is used to compare the h-index of a group of highly cited researchers?". *Australian Journal of Basic and Applied Sciences*, v. 7, n. 4, pp. 198-202. <https://ssrn.com/abstract=2259614>

Gielen, Matt; Rosen, Jeremy (2016). "Reverse engineering the YouTube algorithm: Part I". *Tubefilter.com*, 23 June. <http://www.tubefilter.com/2016/06/23/reverse-engineering-youtube-algorithm>

Giustini, Dean; Boulos, Maged N. K. (2013). "Google Scholar is not enough to be used alone for systematic reviews". *Online journal of public health informatics*, v. 5, n. 2, pp. 1-9. <https://doi.org/10.5210/ojphi.v5i2.4623>

Google (2011). "About Google Scholar". Google Scholar. <http://scholar.google.com/intl/en/scholar/about.html>

Google (2017). "How Google search works. Learn how Google discovers, crawls, and serves web pages". Google. Search console help. <https://support.google.com/webmasters/answer/70897?hl=en>

Harzing, Anne-Wil (2011). *The publish or perish book: your guide to effective and responsible citation analysis*. Melbourne, Australia: Tarma Software Research Pty Ltd. ISBN: 978 1 60752 120 4 <https://harzing.com/publications/publish-or-perish-book/pdf>

Harzing, Anne-Wil (2013). "A preliminary test of Google Schol-

ar as a source for citation data: A longitudinal study of Nobel prize winners". *Scientometrics*, v. 94, n. 3, pp. 1057-1075. <https://doi.org/10.1007/s11192-012-0777-7>

Harzing, Anne-Wil (2014). "A longitudinal study of Google Scholar coverage between 2012 and 2013". *Scientometrics*, v. 98, n. 1, pp. 565-575. https://harzing.com/download/gsc_coverage.pdf <https://doi.org/10.1007/s11192-013-0975-y>

Jacsó, Péter (2008a). "Testing the calculation of a realistic h-index in Google Scholar, Scopus, and Web of Science for FW Lancaster". *Library trends*, v. 56, n. 4, pp. 784-815. <https://goo.gl/Sj5QBr> <https://doi.org/10.1353/lib.0.0011>

Jacsó, Péter (2008b). "The pros and cons of computing the h-index using Google Scholar". *Online information review*, v. 32, n. 3, pp. 437-452. <https://goo.gl/Hz8ABM> <https://doi.org/10.1108/14684520810889718>

Jacsó, Péter (2009). "Calculating the h-index and other bibliometric and scientometric indicators from Google Scholar with the Publish or Perish software". *Online information review*, v. 33, n. 6, pp. 1189-1200. <https://doi.org/10.1108/14684520911011070>

Jacsó, Péter (2012). "Using Google Scholar for journal impact factors and the h-index in nationwide publishing assessments in academia – siren songs and air-raid sirens". *Online information review*, v. 36, n. 3, pp. 462-478. <https://goo.gl/FfrDgt> <https://doi.org/10.1108/14684521211241503>

Jamali, Hamid R.; Asadi, Saeid (2010). "Google and the scholar: the role of Google in scientists' information-seeking behaviour". *Online information review*, v. 34, n. 2, pp. 282-294. <https://goo.gl/cZrNWt> <https://doi.org/10.1108/14684521011036990>

Jamali, Hamid R.; Nabavi, Majid (2015). "Open access and sources of full-text articles in Google Scholar in different subject fields". *Scientometrics*, v. 105, n. 3, pp. 1635-1651. <https://doi.org/10.1007/s11192-015-1642-2>

Lemon, Jim (2006). "Plotrix: a package in the red light district of R". *R-News*, v. 6, n. 4, pp. 8-12. https://www.researchgate.net/publication/260171541_Plotrix_A_package_in_the_red_light_district_of_R

Localseoguide (2016). "Local SEO ranking factors study 2016". Localseoguide. <http://www.localseoguide.com/guides/2016-local-seo-ranking-factors>

Maciá-Domene, Fernando (2015). *SEO: técnicas avanzadas*. Barcelona: Anaya. ISBN: 978 84 41537309

Marcos, Mari-Carmen; González-Caro, Cristina (2010). "Comportamiento de los usuarios en la página de resultados de los buscadores. Un estudio basado en eye tracking". *El profesional de la información*, v. 19, n. 4, pp. 348-358. <https://doi.org/10.3145/epi.2010.jul.03>

Martín-Martín, Alberto; Ayllón, Juan-Manuel; Orduña-Malea, Enrique; Delgado-López-Cózar, Emilio (2016a). *Goog-*

Register for free at <https://www.scipedia.com> to download the version without the watermark

le Scholar Metrics released: a matter of languages... and something else. Granada: Universidad de Granada.
<https://arxiv.org/abs/1607.06260v1>

Martín-Martín, Alberto; Orduña-Malea, Enrique; Ayllón, Juan-Manuel; Delgado-López-Cózar, Emilio (2016b). "Back to the past: On the shoulders of an academic search engine giant". *Scientometrics*, v. 107, n. 3, pp. 1477-1487.
<https://arxiv.org/abs/1603.09111>
<https://doi.org/10.1007/s11192-016-1917-2>

Martín-Martín, Alberto; Orduña-Malea, Enrique; Ayllón, Juan-Manuel; Delgado-López-Cózar, Emilio (2014). *Does Google Scholar contain all highly-cited documents (1950-2013)? EC3 working papers*. Granada: Universidad de Granada.
<https://arxiv.org/abs/1410.8464>

Martín-Martín, Alberto; Orduña-Malea, Enrique; Harzing, Anne-Wil; Delgado-López-Cózar, Emilio (2017). "Can we use Google Scholar to identify highly-cited documents?". *Journal of informetrics*, v. 11, n. 1, pp. 152-163.
<https://arxiv.org/abs/1804.10439>
<https://doi.org/10.1016/j.joi.2016.11.008>

Mayr, Philipp; Walter, Anne-Kathrin (2007). "An exploratory study of Google Scholar". *Online information review*, v. 31, n. 6, pp. 814-830.
<https://arxiv.org/pdf/0707.3575.pdf>
<https://doi.org/10.1108/14684520710841784>

Merton, Robert K. (1968). "The Matthew effect in science". *Science*, v. 159, n. 3810, pp. 56-63.
<https://goo.gl/Zcpqcc>
<https://doi.org/10.1126/science.159.3810.56>

Moed, Henk F.; Bar-Ilan, Judit; Halevi, Gali (2016). "A new methodology for comparing Google Scholar and Scopus". *Journal of informetrics*, v. 10, n. 2, pp. 533-551.
<https://arxiv.org/abs/1512.05741>
<https://doi.org/10.1016/j.joi.2016.04.017>

MOZ (2015). *Search engine ranking factors 2015*.
<https://moz.com/search-ranking-factors/correlations>

MOZ (2017). *Google algorithm change history*.
<https://moz.com/google-algorithm-change>

Muñoz-Martín, Beatriz (2015). "Incrementa el impacto de tus artículos y blogs: de la invisibilidad a la visibilidad". *Revista de la Sociedad Otorrinolaringológica de Castilla y León, Cantabria y La Rioja*, v. 6, n. Suppl. 4, pp. 6-32.
<http://hdl.handle.net/10366/126907>

Orduña-Malea, Enrique; Ayllón, Juan-Manuel; Martín-Martín, Alberto; Delgado-López-Cózar, Emilio (2014). *About the size of Google Scholar: playing the numbers. EC3 working papers*. Granada: Universidad de Granada.
<https://arxiv.org/abs/1407.6239>

Orduña-Malea, Enrique; Ayllón, Juan-Manuel; Martín-Martín, Alberto; Delgado-López-Cózar, Emilio (2015). "Methods for estimating the size of Google Scholar". *Scientometrics*, v. 104, n. 3, pp. 931-949.
<https://doi.org/10.1007/s11192-015-1614-6>

Orduña-Malea, Enrique; Martín-Martín, Alberto; Ayllón, Juan-Manuel; Delgado-López-Cózar, Emilio (2016). *La revolución Google Scholar: destapando la caja de Pandora académica*. Granada: Editorial Universidad de Granada. ISBN: 978 84 33859419
<https://goo.gl/3oUGKQ>

Pedersen, Lee A.; Arendt, Julie (2014). "Decrease in free computer science papers found through Google Scholar". *Online information review*, v. 38, n. 3, pp. 348-361.
<https://goo.gl/ngmZ1Q>
<https://doi.org/10.1108/OIR-07-2013-0159>

R Development Core Team (2008). *R: a language and environment for statistical computing*.
<http://softlibre.unizar.es/manuales/aplicaciones/r/fullrefman.pdf>
<https://www.R-project.org>

Ratcliff, Christopher (2016). "WebPromo's Q&A with Google's Andrey Lipattsev, search engine watch". *Search engine watch*, 24 March.
<https://searchenginewatch.com/2016/04/06/webpromos-qa-with-googles-andrey-lipattsev-transcript>

Revelle, William (2017). *Psych: procedures for personality and psychological research*. Northwestern University.
<https://CRAN.R-project.org/package=psych>

Schwartz, Barry (2016). "Now we know: Here are Google's top 3 search ranking factors". *Search engine land*, 24 March.
<http://searchengineland.com/now-know-googles-top-three-search-ranking-factors-245882>

Searchmetrics (2016). "Rebooting ranking factors". *Searchmetrics*.
<http://www.searchmetrics.com/knowledge-base/ranking-factors>

Torres-Salinas, Daniel; Ruiz-Pérez, Rafael; Delgado-López-Cózar, Emilio (2009). "Google scholar como herramienta para la evaluación científica". *El profesional de la información*, v. 18, n. 5, pp. 501-510.
<https://doi.org/10.3145/epi.2009.sep.03>

Van-Aalst, Jan (2010). "Using Google Scholar to estimate the impact of journal articles in education". *Educational researcher*, v. 39, n. 5, pp. 387-400.
<https://goo.gl/p1mDBi>

Van-der-Graaf, Peter (2012). "Reverse engineering search engine algorithms is getting harder". *Search engine watch*, 7 June.
<https://searchenginewatch.com/sew/how-to/2182553/reverse-engineering-search-engine-algorithms-getting-harder>

Walters, William H. (2008). "Google Scholar search performance: Comparative recall and precision". *Portal: Libraries and the academy*, v. 9, n. 1, pp. 5-24.
<https://goo.gl/9aw2fd>
<https://doi.org/10.1353/pla.0.0034>

Wiley (2015). *Writing for SEO*.
<https://authorservices.wiley.com/author-resources/Journal-Authors/Prepare/writing-for-seo.html>